Diabetologia (2011) 54:1702–1709
DOI 10.1007/s00125-011-2161-1

ARTICLE

# The *HLA-B*3906* allele imparts a high risk of diabetes only on specific *HLA-DR/DQ* haplotypes

**E. E. Baschal · P. R. Baker · K. R. Eyring · J. C. Siebert · J. M. Jasinski · G. S. Eisenbarth**

Received: 4 August 2010 / Accepted: 29 March 2011 / Published online: 1 May 2011


**Abstract**
*Aims/hypothesis* We investigated the risk associated with *HLA-B*39* alleles in the context of specific *HLA-DR/DQ* haplotypes.
*Methods* We studied a readily available dataset from the Type 1 Diabetes Genetics Consortium that consists of 2,300 affected sibling pair families genotyped for both HLA alleles and 2,837 single nucleotide polymorphisms across the major histocompatibility complex region.
*Results* The *B*3906* allele significantly enhanced the risk of type 1 diabetes when present on specific *HLA-DR/DQ* haplotypes (*DRB1*0801-DQB1*0402*: $p=1.6\times10^{-6}$, OR 25.4; *DRB1*0101-DQB1*0501*: $p=4.9\times10^{-5}$, OR 10.3) but did not enhance the risk of *DRB1*0401-DQB1*0302* haplotypes. In addition, the *B*3901* allele enhanced risk on the *DRB1*1601-DQB1*0502* haplotype ($p=3.7\times10^{-3}$, OR 7.2).
*Conclusions/interpretation* These associations indicate that the *B*39* alleles significantly increase risk when present on specific *HLA-DR/DQ* haplotypes, and HLA-B typing in concert with specific *HLA-DR/DQ* genotypes should facilitate genetic prediction of type 1 diabetes, particularly in a research setting.

**Keywords** Association study · Genetics · HLA class I · HLA-B · *HLA-B*39* · MHC · Type 1 diabetes

**Abbreviations**

| | |
|---|---|
| AFBAC | Affected family based control |
| IMGT | International ImMunoGeneTics information system |
| SNP | Single-nucleotide polymorphism |
| T1DGC | Type 1 Diabetes Genetics Consortium |
| TDT | Transmission disequilibrium test |
| WTCCC | Wellcome Trust Case Consortium |

## Introduction

Alleles of multiple genes within the MHC region are associated with type 1 diabetes. In particular, the *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1* genes are most strongly associated with type 1 diabetes. Specifically, the common genotype with the highest risk for type 1 diabetes in Europeans is *DRB1*0301-DQA1*0501-DQB1*0201/DRB1*04-DQA1*0301-DQB1*0302* (DR3/4), which is present in 2.4% of the general US population [1, 2]. Class I HLA alleles have also been associated with type 1 diabetes, particularly the *HLA-B*39* and *HLA-A*24* alleles [3–6]. The high risk of *HLA-B*3906* has been documented for *DRB1*0101-DQB1*0501* and *DRB1*0801-DQB1*0402* chromosomes [4], and the *B**3906 allele alone has been associated with a younger age of onset of type 1 diabetes [4, 5, 7, 8]. The risk of the *B*3901* allele has not been well documented.

E. E. Baschal and P. R. Baker contributed equally to this study.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00125-011-2161-1) contains supplementary material, which is available to authorised users.

E. E. Baschal · P. R. Baker · K. R. Eyring · J. M. Jasinski · G. S. Eisenbarth (✉)
Barbara Davis Center for Childhood Diabetes,
University of Colorado Denver,
1775 Aurora Ct, Box B140, Aurora, CO 80045-6511, USA
e-mail: George.Eisenbarth@ucdenver.edu

J. C. Siebert
CytoAnalytics,
Denver, CO, USA

*B*3906* and *B*3901* are the most common *HLA-B*39* alleles. As we are interested in extended haplotypes that span *HLA-DR/DQ* and *HLA-B* (both class I and class II genes), in this paper we investigate associations of risk for type 1 diabetes for multiple *HLA-DR/DQ* haplotype groups with either the *B*3906* or *B*3901* allele. We show that there is increased risk for both *B*3906* and *B*3901* on specific *HLA-DR/DQ* haplotypes, and extend the analysis to examine the single-nucleotide polymorphism (SNP)-level variation of these haplotypes.

## Methods

### Study populations and genotyping

This analysis included 2,300 affected sibling-pair families (10,012 individuals typed for HLA and/or SNPs) from the Type 1 Diabetes Genetics Consortium (T1DGC), using the 2007.11.MHC data freeze [9]. Affected sibling pairs and their parents were enrolled in nine cohorts worldwide. Within the analysed cohorts of Asia-Pacific, Europe, North America, UK, British Diabetes Association (BDA), Danish, Human Biological Data Interchange (HBDI), Joslin and Sardinian, 99% of individuals are classified as white or unknown. The T1DGC performed basic quality control analyses on the data. All study participants or their parents/surrogates provided written informed consent to participate, and the study protocol was approved by the relevant Ethics Committees and Institutional Review Boards.

Genotyping was completed for 3,072 SNPs at the Wellcome Trust Sanger Institute, using custom Illumina exon-centric and mapping panels (2957 distinct SNPs [1536 SNPs in each panel with 115 overlapping SNPs] with 2837 of 2957 SNPs successfully typed, yielding a 96% SNP success rate). In addition, complete four-digit HLA typing (*HLA-DPB1*, *HLA-DPA1*, *HLA-DQB1*, *HLA-DQA1*, *HLA-DRB1*, *HLA-B*, *HLA-C* and *HLA-A*), performed using immobilised probe linear arrays, was available for all samples [10].

*MICA genotyping* To analyse association of *HLA-B*39* with *MICA* alleles, we typed 341 T1DGC individuals who had a *B*39* allele or were family members of individuals with a *DRB1*0404* allele. *MICA* genotypes were determined using a fluorescence-based method as reported previously by Zake et al., Park et al. and Triolo et al. [11–13]. Details of the exact method used by our laboratory have been reported previously by Triolo et al. [13].

### Data processing

SNP positions used NCBI Build 36 (National Center for Biotechnology Information, U.S. National Library of Medicine, Bethesda, MD, USA). T1DGC chromosomes were generated from SNP and HLA genotype data using multiple software packages. First, to establish that the genotype data demonstrated a Mendelian inheritance pattern within each family, the PedCheck program (http://watson.hgen.pitt.edu) [14] was used on data from both Illumina panels and HLA separately. Mendelian inheritance patterns were present for all families. Next, data from the Illumina mapping SNP panel, the exon-centric SNP panel, and HLA were combined using custom Java programs. Merlin software (www.sph.umich.edu/csg/abecasis/Merlin) [15] was used to phase the SNP and HLA genotype data from families into chromosomes. In situations of ambiguous phase (resulting from heterozygous SNPs or HLA in all family members), phase was not inferred and instead unphased alleles were labelled as such and were excluded from analyses where appropriate.

Founder chromosomes were used in these analyses, yielding four unique chromosomes per family. An affected family based control (AFBAC) method was used to assign case or control status to chromosomes using Microsoft Excel macros as previously described [16–19]. Founder chromosomes were labelled as case if they were ever transmitted to a case individual. Correspondingly, chromosomes were labelled as control if they were never transmitted to a case individual. Two affected children in each family were used to label the chromosomes as either case or control, as this was the ascertainment scheme for T1DGC. Analysis of founder chromosomes avoids the issues of non-independence of siblings in the family, as only the four parental (founder) chromosomes are analysed in every family.

Amino acid sequences for *HLA-B*3906*, *B*3901* and *B*1402* were taken from the international ImMunoGeneTics information system (IMGT) database (www.ebi.ac.uk/imgt/hla/align.html). Three-dimensional modelling of the HLA-B peptide binding region was performed using IMGT (http://imgt.cines.fr/3Dstructure-DB/) [20].

### Statistical analysis

The Fisher's exact test (two-sided) was used to calculate $p$ values for AFBAC founder chromosome association with type 1A diabetes, with $\alpha=0.05$.

Each founder chromosome was classified as transmitted or not transmitted based on transmission to one randomly selected case child in each family. A modified transmission disequilibrium test (TDT) was used to evaluate the significance for each *HLA-DR/DQ* group by comparing, within each *HLA-DR/DQ* group, the number of *B*3906* alleles transmitted to an affected individual, the number of *B*3906* alleles not transmitted to an affected individual, the number of non-*B*3906* alleles transmitted to an affected

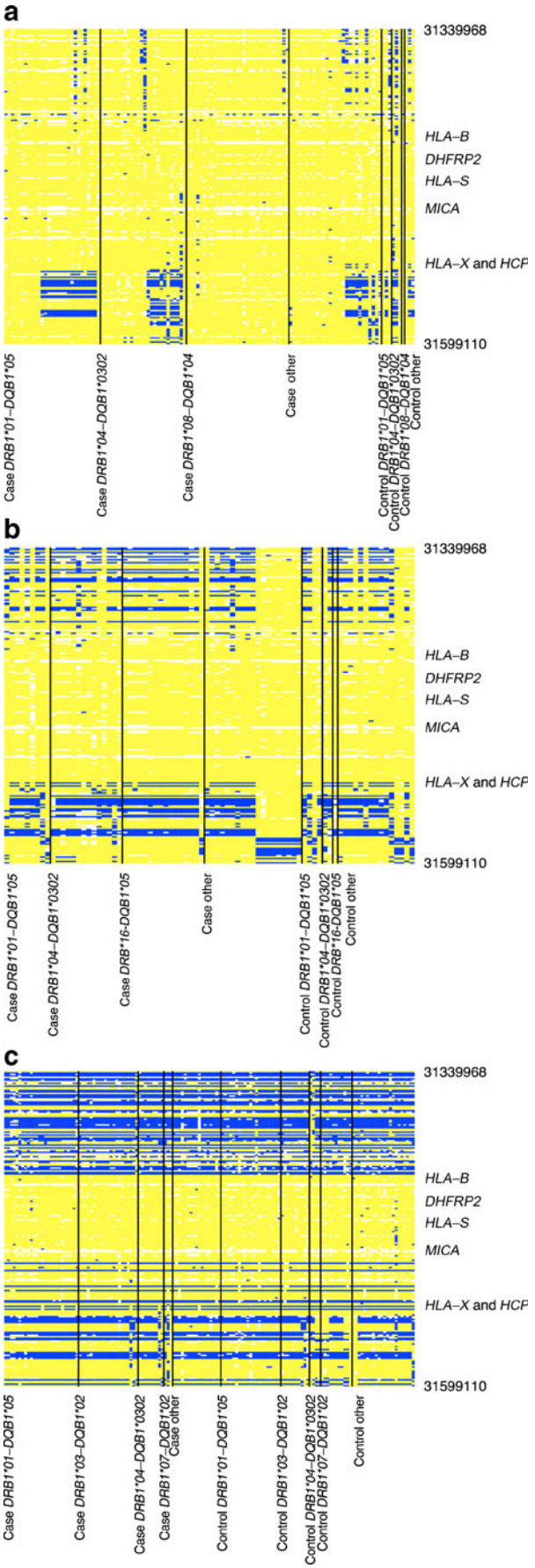


**Fig. 1** Comparison of *HLA-B* alleles based on SNPs. Each column is a chromosome and each row is a SNP. **a** *HLA-B*3906* chromosomes compared with *HLA-B*3906* consensus (124 chromosomes, 219 SNPs and 0.26 Mb). **b** *HLA-B*3901* chromosomes compared with *HLA-B*3906* consensus (80 chromosomes, 219 SNPs and 0.26 Mb). **c** *HLA-B*1402* chromosomes compared with *HLA-B*3906* consensus (144 chromosomes, 219 SNPs and 0.26 Mb). The same consensus sequence was used in **a**–**c**, and was based on *B*3906* chromosomes. The chromosomes are organised by *HLA-DR/DQ* type in each figure part, and only the region surrounding the *HLA-B* gene is shown. Yellow boxes show that the allele matches that of the *B*3906* consensus, whereas blue represents the opposite allele. The telomeric end of the MHC is at the top of the graph. Type 1 diabetes case chromosomes are on the left and control chromosomes are on the right

individual and the number of non-*B*3906* alleles not transmitted to an affected individual. These values were compared using the Fisher's exact test ($\alpha$=0.05). The same logic was used for *B*3901*. The underlying counts were computed using custom Java programs, which are available upon request.

Relative risk for the *DRB1*08-B*39/DRB1*03* genotype was analysed using case T1DGC individuals (one case per family) and calculated control-genotype frequencies taken from non-transmitted control T1DGC chromosomes. We required the *B*39* allele to be present on the *DRB1*08* chromosome, not the *DRB1*03* chromosome. Relative risk was calculated using Graph Pad 4.0, and the absolute risk estimate was based on a background population risk of 1/300.

For Fig. 1 and electronic supplementary material (ESM) Figs 1–3, consensus sequences were identified using an automated version of the process previously described for evaluating extended haplotypes [21]. Briefly, given a haplotype matrix consisting of a range of contiguous SNPs and a specific number of chromosomes, we identified the most common sequence of SNPs over a small window or subrange of, for example, 30 SNPs. Then, we narrowed the window to a smaller number of SNPs (e.g. ten) and repeated the process, thereby deriving a consensus sequence for the entire range. Additionally, as the sequence emerged, we scored each chromosome for identity with the emerging sequence. Chromosomes that did not sufficiently match the consensus were eliminated from further consideration. A consensus sequence was derived for the *B*3906* chromosomes in the region surrounding the *HLA-B* gene, and then the other groups of chromosomes were compared with this *B*3906* consensus in Fig. 1 and ESM Figs 1–3.

The number of chromosomes differs between Fig. 1 and the ESM figures because of the logic that excludes a chromosome from congruence analysis if more than 20% of the SNPs within the range of consideration are unknown or unphased. As Fig. 1 and the ESM figures consider different ranges, different chromosomes fail across those ranges.

## Results

We interrogated a readily available dataset from the T1DGC to examine the risk associated with the two most common *HLA-B*39* alleles, *B*3906* and *B*3901*. The *B*3906* allele is present in 2.6% of case and 0.4% of control chromosomes, while *B*3901* is present in 1.3% of case and 1.0% of control chromosomes. Using case and control chromosomes classified using an AFBAC method, the *B*3906* allele is associated with very high risk for type 1 diabetes (127/141 *B*3906* [90% case chromosomes] vs 4963/7650 non-*B*3906* [65% case], $p=1.1\times10^{-11}$, OR 4.9). These results combine all *HLA-DR/DQ* haplotypes together and look at the overall risk of each allele only. Overall, the *B*3901* allele was not significantly associated with type 1 diabetes (68/95 *B*3901* [72% case] vs 5022/7696 non-*B*3901* [65% case] $p=0.2$, OR 1.3). We next investigated the associations of specific *HLA-DR/DQ* haplotypes with *B*3901* and *B*3906* (Table 1). The great majority of *B*3901* and *B*3906* alleles are concentrated on relatively few *HLA-DR/DQ* haplotypes. Specifically, *B*3906* is present almost exclusively on *DRB1*0101-DQB1*0501*, *DRB1*0801-DQB1*0402*, and *DRB1*0401-DQB1*0302* haplotypes, whereas *B*3901* is present on *DRB1*0101-DQB1*0501* and *DRB1*1601-DQB1*0502* haplotypes.

We then assessed the risk of both *B*39* alleles on these stratified *HLA-DR/DQ* haplotypes compared with the results presented earlier which are not stratified by *HLA-DR/DQ* haplotype. *B*3906* is associated with case chromosomes on *DRB1*0801-DQB1*0402* ($p=1.6\times10^{-6}$, OR 25.4, Table 2) and is also associated on *DRB1*0101-DQB1*0501* haplotypes ($p=4.9\times10^{-5}$, OR 10.3). There is a trend toward increased association of the *HLA-B*3906* allele on other *HLA-DR/DQ* haplotypes, with the exception of *DRB1*0401-DQB1*0302* where there is no increased risk associated with the *B*3906* allele ($p=0.7$, OR 0.8). When we examine the *B*3901* allele, we find that it is associated with case chromosomes on the *DRB1*1601-DQB1*0502* haplotype ($p=3.7\times10^{-3}$, OR 7.2, Table 3). *B*3901* was not significantly associated with diabetes on other *HLA-DR/DQ* haplotypes. Analysis of transmission of these *B*39*-bearing haplotypes to one affected child per family is concordant with the above case/control analyses (Tables 4 and 5). Both *B*3906* and *B*3901* significantly enhance type 1 diabetes risk, each on specific *HLA-DR/DQ* haplotypes.

**Table 1** Distribution of *B*3901* and *B*3906* alleles on *HLA-DR/DQ* haplotypes

| Haplotype | | Total | *B*3906* | | *B*3901* | |
|---|---|---|---|---|---|---|
| *DRB1* | *DQB1* | | *n* | % of DR | *n* | % of DR |
| *0101 | *0501 | 477 | 29 | 6.08 | 11 | 2.31 |
| *0801 | *0402 | 153 | 37 | 24.18 | 4 | 2.61 |
| *0401 | *0302 | 1,153 | 19 | 1.65 | 8 | 0.69 |
| *0301 | *0201 | 1,674 | 10 | 0.60 | 8 | 0.48 |
| *1601 | *0502 | 131 | 5 | 3.82 | 17 | 12.98 |
| Other | – | 2,323 | 26 | 1.12 | 32 | 1.38 |

*HLA-DR/DQ* groups are shown in the table if there were at least ten chromosomes with either *B*3906* or *B*3901*. *B*3906* and *B*3901* are present on relatively few *HLA-DR/DQ* haplotypes.

The 'other' group consists of all *HLA-DR/DQ* haplotypes that are not specified in the table

The *B*3906* and *B**3901 amino acid sequences are closely related, and their nearest non-*B**39 *HLA-B* allele is *B*1402* [22]. We decided to investigate these three alleles in more detail. Only two amino acids differ between the *B*3906* and *B*3901* alleles, both in the base of the peptide-binding pocket (*B*3906* to *B*3901*: threonine to arginine and tryptophan to leucine). When we compare *B*1402* and *B*3906*, they differ by seven amino acids, five of which appear to be in the peptide-binding region. However, despite this apparent similarity at the amino acid sequence level, the *B*1402* allele does not enhance diabetes risk (82/163 *B*1402* [50% case] vs 5008/7628 non-*B*1402* [66% case], $p=8.2E{-}5$, OR 0.5). *B*1402* is marginally associated with decreased type 1 diabetes risk on *DRB1*0701-DQB1*0201* and *DRB1*0301-DQB1*0201* haplotypes (ESM Table 1).

We analysed SNPs surrounding the *HLA-B* gene for the three *HLA-B* alleles of interest to determine the SNP-level differences (Fig. 1a–c). In these figures, each column is one chromosome (panels showing *B*3906*, *B*3901* or *B*1402* chromosomes, respectively) and each row is one SNP. These figures only show the region surrounding the *HLA-B* gene (full chromosome plots can be found in ESM Figs 1–3). All three groups of chromosomes are compared with a common consensus sequence for *B*3906*. Yellow boxes represent SNP alleles that match the consensus while blue represents SNP alleles that do not match the consensus. All three groups of chromosomes (*B*3906*, *B*3901* and *B*1402*) are congruent (or match) for a region of SNPs surrounding *HLA-B* (59,530 base pairs), even though *B*3906* and *B*3901* are associated with higher risk and *B*1402* is not. *B*3901* and *B*3906* chromosomes are congruent (based on SNPs) with each other for 142,861 base pairs surrounding *HLA-B*. We typed a subset of families for the *MICA* gene, which is in the region of SNP identity, enriching for families with *B*39*. Of the individuals (a mixture of patients and unaffected individuals from T1DGC families) with the *B*39* allele, 100% had the *MICA* 9 allele (26/26) vs 17% of individuals without *B*39* (54/314, $p=1.3\times10^{-18}$).

## Discussion

*HLA-B*39* risk has been well defined in the literature [3, 4, 7, 8]. With the characterisation of conserved extended

**Table 2** The *HLA-B*3906* allele is significantly associated with type 1 diabetes on specific *HLA-DR/DQ* haplotypes

| Haplotype | | Case | | | Control | | | OR (95% CI) | *p* value |
|---|---|---|---|---|---|---|---|---|---|
| *DRB1* | *DQB1* | With *B*3906* | Without *B*3906* | % with *B*3906* | With *B*3906* | Without *B*3906* | % with *B*3906* | | |
| *801 | *402 | 36 | 68 | 35 | 1 | 48 | 2.0 | 25.4 (3.37–191.8) | $1.6\times10^{-6}$ |
| *101 | *501 | 27 | 254 | 10 | 2 | 194 | 1.0 | 10.31 (2.42–43.89) | $4.9\times10^{-5}$ |
| *401 | *302 | 17 | 1034 | 1.6 | 2 | 100 | 2.0 | 0.82 (0.19–3.61) | 0.68 |
| *301 | *201 | 10 | 1365 | 0.7 | 0 | 299 | 0.0 | 4.61 (0.27–78.82) | 0.22 |
| Other | – | 26 | 1458 | 1.8 | 5 | 1423 | 0.4 | 5.19 (1.94–13.25) | $1.9\times10^{-4}$ |

Analysis of T1DGC AFBAC case and control chromosomes for the risk associated with *B*3906*, stratified by *HLA-DR/DQ* haplotype

*B*3906 is associated with case chromosomes on *DRB1*0801-DQB1*0402* ($p=1.6\times10^{-6}$, OR 25.4) and *DRB1*0101-DQB1*0501* haplotypes ($p=4.9\times10^{-5}$, OR 10.3)

The 'other' group consists of all *HLA-DR/DQ* haplotypes that are not specified in the table

haplotypes in type 1 diabetes risk [18, 23], studies exploring the genetic context of *HLA-B*39* alleles (via haplotype and SNP analysis) have become increasingly important in further characterising haplotypic risk for type 1 diabetes and those MHC regions involved in risk determination [4, 24, 25]. Here we confirm previously discovered haplotypic associations of *HLA-B*39* alleles, and place these associations in the context of high-density SNP analysis.

*B*3906* is associated with increased risk on *DRB1*0801-DQB1*0402* and *DRB1*0101-DQB1*0501* haplotypes, whereas *B*3901* is associated with increased risk on *DRB1*1601-DQB1*0502* haplotypes. While both *B*39* alleles are associated with increased risk, *B*1402*, their nearest non-*B*39 HLA-B* allele by amino acid sequence, is not associated with increased risk [22]. Even one amino acid difference, particularly in the peptide-binding region, could influence disease risk associated with the allele. Therefore, even though there are only seven amino acid differences (five in the peptide-binding region) between *B*1402* and *B*3906*, those changes could result in the drastic difference in risk seen in this analysis. In addition, as there are two amino acid differences between *B*3906* and *B*3901*, both in the peptide-binding region, those differences could also result in differences in disease risk. Therefore, we hypothesise that *B*39*-associated increased risk is due to the differences in amino acid sequence between *B*3901*/*B*3906* and *B*1402* or that the risk is associated with polymorphisms where *B*3906* and *B*3901* are congruent for almost all SNPs but do not match the *B*1402* SNP haplotype. Gene-conversion events between alleles of the *HLA-B* gene have been reported previously. This is one possible explanation for the presence of highly similar SNP sequences surrounding the *HLA-B* gene for the *B*3906*, *B*3901* and *B*1402* alleles [26–28].

There are five genes in the 142,861 base pair region of near SNP identity between the *B*3906* and *B*3901* haplotypes: *HLA-B*, *DHFRP2* (dihydrofolate reductase pseudogene 2), *HLA-S* (pseudogene), *MICA* (MHC class I polypeptide-related sequence A), and *HLA-X* (pseudogene). Theoretically, any of these genes or even variations in non-coding sequence in the region could contribute to the increased risk associated with *B*3901* and *B*3906*.

**Table 3** The *HLA-B*3901* allele is significantly associated with type 1 diabetes on specific *HLA-DR/DQ* haplotypes

| Haplotype | | Case | | | Control | | | OR (95% CI) | *p* value |
|---|---|---|---|---|---|---|---|---|---|
| *DRB1* | *DQB1* | With *B*3901* | Without *B*3901* | % with *B*3901* | With *B*3901* | Without *B*3901* | % with B*3901 | | |
| *1601 | *502 | 15 | 58 | 21 | 2 | 56 | 3.4 | 7.24 (1.58–33.12) | $3.68\times10^{-3}$ |
| *101 | *501 | 8 | 273 | 2.8 | 3 | 193 | 1.5 | 1.89(0.49–7.2) | 0.54 |
| Other | – | 34 | 3774 | 0.9 | 18 | 1785 | 1.0 | 0.89(0.5–1.59) | 0.77 |

Analysis of T1DGC AFBAC case and control chromosomes for the risk associated with *B*3901*, stratified by *HLA-DR/DQ* haplotype

*B*3901* is associated with case chromosomes on the *DRB1*1601-DQB1*0502* haplotype ($p=3.7\times10^{-3}$, OR 7.2)

The 'other' group consists of all *HLA-DR/DQ* haplotypes that are not specified in the table

**Table 4** Transmission of the *B*3906* allele

| Haplotype | | *B*3906* | | Not *B*3906* | | *p* value |
|---|---|---|---|---|---|---|
| *DRB1* | *DQB1* | Transmitted to affected | Not transmitted to affected | Transmitted to affected | Not transmitted to affected | |
| *0101 | *0501 | 23 | 6 | 190 | 255 | $1.5\times10^{-4}$ |
| *0301 | *0201 | 10 | 0 | 1,111 | 546 | 0.04 |
| *0401 | *0302 | 14 | 3 | 877 | 253 | 0.8 |
| *0404 | *0302 | 8 | 1 | 197 | 105 | 0.2 |
| *0801 | *0402 | 29 | 8 | 50 | 65 | $2.7\times10^{-4}$ |
| Other (≤5) | – | 16 | 6 | 582 | 1,201 | 0.2 |

TDT analysis was completed to calculate the transmission of *B*3906* to one case individual per family, stratified by *HLA-DR/DQ* alleles

*HLA-DR/DQ* groups with five or fewer *B*3906* alleles are combined into an 'other' category

There are limitations to this work. First, *B*39* is not a common allele, as we see it in only 3% of individuals with type 1 diabetes in this dataset. Consequently, we cannot tell whether *B*3906* haplotypes confer greater risk than *B*3901* haplotypes as the only *HLA-DR/DQ* haplotype that is common to both alleles is *DRB1*0101-DQB1*0501*, and the numbers in this group are too small to distinguish the risk associated with the alleles. Therefore, a larger dataset is needed to fully differentiate the risk associated with these two *B*39* alleles. It has been previously described that the *B*3906* allele is high risk on the *DRB1*0404-DQB1*0302* haplotype [8, 29, 30]. Our results are in the same direction, but a larger dataset would be needed to confirm the effect (*DRB1*0404-DQB1*0302*: 9/9 *B*3906* [100% of *B*3906* are case chromosomes] compared with 249/304 non-*B*3906* [82% of non-*B*3906* are case chromosomes], $p$=0.4, OR 4.2).

We did not observe a difference in risk due to the *B*3906* allele on the *DRB1*0401-DQB1*0302* haplotype (*DRB1*0401-DQB1*0302*: 17/19 *B*3906* [89% of *B*3906* are case chromosomes] compared with 1034/1134 non-*B*3906* [91% of non-*B*3906* are case chromosomes], $p$=0.7, OR 0.82). It is possible that the *DRB1*0401-DQB1*0302* risk is so high that we are unable to see an incremental effect from *B*3906*. In addition, as the numbers in the *B*3906-DRB1*0401-DQB1*0302* group are small, it is possible that the study was underpowered to find an association. Similarly, the *B*3901* allele is rare (1.2%). The current study should be considered hypothesis-generating given the limited number of chromosomes containing *B*39*, even in this large T1DGC dataset.

To replicate our findings from the T1DGC dataset we employed a large, publicly available dataset from the Wellcome Trust Case Consortium (WTCCC) [5] (ESM Table 2). Diplotypes containing high-risk *DRB1-DQB1* haplotypes (namely *DRB1*0801-DQB1*0402*, *DRB1*0101-DQB1*0501*, *DRB1*301-DQB1*0302*, *DRB1*0401-DQB1*0302*, and *DRB1*1601-DQB1*0502*) were evaluated for the presence of *HLA-B*3906* and *HLA-B*3901*. Those containing *DRB1*0801-DQB1*0402* (but not the other high-risk haplotypes) were more likely to have *HLA-B*3906* (case 42% [8/19], control 6% [3/53], $p=6.3\times10^{-4}$, OR 12). Similar results were found for genotypes containing *DRB1*0101-DQB1*0501* (case 9% [7/81], control 2% [7/367], $p=5.8\times10^{-3}$, OR 5) and *DRB1*0301-DQB1*0201* (case 5% [25/502], control 1% [9/628], $p=6.7\times10^{-4}$, OR 4). Our finding that the haplotype *DRB1*0401-DQB1*0302* is not associated with *HLA-B*3906* was also confirmed, with similar frequency between cases (3% [5/181]) and controls (2% [3/186], $p$=0.5).

Our analyses using both the T1DGC and WTCCC datasets, therefore, indicate that *HLA-B*3906* is high risk on specific *DRB1-DQB1* haplotypes. At a practical level, identifying specific *HLA-DR/DQ* haplotypes also carrying *B*39* will aid type 1 diabetes genetic prediction for individuals, particularly in a research setting. In terms of quantificative risk and application in a clinical setting,

**Table 5** Transmission of the *B*3901* allele

| Haplotype | | *B*3901* | | Not *B*3901* | | *p* value |
|---|---|---|---|---|---|---|
| *DRB1* | *DQB1* | Transmitted to affected | Not transmitted to affected | Not transmitted to affected | Not transmitted to affected | |
| *0101 | *0501 | 6 | 4 | 207 | 257 | 0.4 |
| *0301 | *0201 | 6 | 2 | 1,115 | 544 | 1 |
| *0401 | *0302 | 6 | 2 | 885 | 254 | 1 |
| *0404 | *0302 | 6 | 0 | 199 | 106 | 0.1 |
| *1601 | *0502 | 12 | 5 | 42 | 72 | 0.02 |
| Other (≤5) | – | 10 | 20 | 511 | 1,133 | 0.8 |

TDT analysis was completed to calculate the transmission of *B*3901* to one case individual per family, stratified by *HLA-DR/DQ* alleles

*HLA-DR/DQ* groups with five or fewer *B*3901* alleles are combined into an 'other' category

genotypes (i.e. diplotypes or both HLA haplotypes) are critical determinates, especially such high-risk diplotypes as DR3/4 with *HLA-B*3906*. In that the *B*39* alleles are uncommon among patients with type 1 diabetes, identifying *B*39* alleles will have little impact on the prediction of overall risk of type 1 diabetes (as illustrated by an ROC curve). In particular, absence of a *B*39* allele would not appreciably decrease the overall risk of type 1 diabetes. In contrast the very high odds ratios of *B*39* in combination with specific *HLA-DR/DQ* alleles (e.g. *DRB1*0801-DQB1*0402*) indicate that for individuals carrying these genotypes there would be a potentially important increase in risk with the presence of *B*39*. Absolute diabetes risk of genotype *DRB1*08-B*39/DRB1*03* is estimated to be 5%, compared with 0.26% when *B*39* is not present with this same genotype. This 5% risk for the *DRB1*08-B*39/DRB1*03* genotype is comparable with the risk seen with the highest risk genotype DR3/4.

From a theoretical point of view, if the effects of *B*39* alleles are dependent on both specific *HLA-DR/DQ* haplotypes and in particular the specific *B*39* sequences, it suggests a potentially complex model of pathogenesis. The leading hypothesis is that *B*39* alleles present specific islet peptides to CD8 T lymphocytes. *HLA-DR/DQ* alleles might determine CD4 T cells targeting of specific auto-antigens and *B*39* alleles would then enhance CD8 T cells targeting of peptides of these autoantigens. Alternatively it is possible that class I polymorphisms may even influence thymic selection and the repertoire of CD4 T cell receptors. Understanding the mechanism by which *B*39* alleles enhance diabetes risk on specific *HLA-DR/DQ* haplotypes will likely contribute to our overall understanding of the loss of tolerance leading to type 1 diabetes.

**Acknowledgements** This research used resources provided by the T1DGC, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD) and the Juvenile Diabetes Research Foundation International (JDRF). This work was supported by the National Institutes of Health (R01 DK32083), Diabetes Autoimmunity Study in the Young (DAISY, R37 DK32493), Autoimmunity Prevention Center (AI050864), Diabetes Endocrine Research Center (P30 DK57516), the Immune Tolerance Network (AI15416), the American Diabetes Association, the Juvenile Diabetes Research Foundation, the Children's Diabetes Foundation and the Brehm Coalition. We would like to thank R. Wong and C. S. Toews of the DERC Molecular Biology Core at the University of Colorado (Denver, CO, USA) for *MICA* genotyping. The DERC Molecular Biology Core is supported by NIH P30 DK57516. J. Siebert is a founder and shareholder of CytoAnalytics. P. R. Baker is a Fellow of the Pediatric Scientist Development Program. The project described was supported by Award K12-HD000850 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

**Duality of interest** The authors declare that there is no duality of interest associated with this manuscript.

## References

1. She JX, Marron MP (1998) Genetic susceptibility factors in type 1 diabetes: linkage, disequilibrium and functional analyses. Curr Opin Immunol 10:682–689
2. Rewers M, Bugawan TL, Norris JM et al (1996) Newborn screening for HLA markers associated with IDDM: Diabetes Autoimmunity Study in the Young (DAISY). Diabetologia 39:807–812
3. Noble JA, Valdes AM, Bugawan TL, Apple RJ, Thomson G, Erlich HA (2002) The HLA class I A locus affects susceptibility to type 1 diabetes. Hum Immunol 63:657–664
4. Valdes AM, Erlich HA, Noble JA (2005) Human leukocyte antigen class I B and C loci contribute to type 1 diabetes (T1D) susceptibility and age T1D onset. Hum Immunol 66:301–313
5. Nejentsev S, Howson JM, Walker NM et al (2007) Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature 450:887–892
6. Howson JM, Walker NM, Clayton D, Todd JA (2009) Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. Diabetes Obes Metab 11(Suppl 1):31–45
7. Howson JM, Walker NM, Smyth DJ, Todd JA (2009) Analysis of 19 genes for association with type I diabetes in the Type I Diabetes Genetics Consortium families. Genes Immun 10(Suppl 1): S74–S84
8. Nejentsev S, Gombos Z, Laine AP et al (2000) Non-class II HLA gene associated with type 1 diabetes maps to the 240-kb region near HLA-B. Diabetes 49:2217–2221
9. Brown WM, Pierce J, Hilner JE et al (2009) Overview of the MHC fine mapping data. Diabetes Obes Metab 11(Suppl 1):2–7
10. Mychaleckyj JC, Noble JA, Moonsamy PV et al (2010) HLA genotyping in the international Type 1 Diabetes Genetics Consortium. Clin Trials 7:S75–87
11. Zake LN, Ghaderi M, Park YS, Babu S, Eisenbarth G, Sanjeevi CB (2002) MHC class I chain-related gene alleles 5 and 5.1 are transmitted more frequently to type 1 diabetes offspring in HBDI families. Ann NY Acad Sci 958:309–311
12. Park YS, Sanjeevi CB, Robles D et al (2002) Additional association of intra-MHC genes, MICA and D6S273, with Addison's disease. Tissue Antigens 60:155–163
13. Triolo TM, Baschal EE, Armstrong TK et al (2009) Homozygosity of the polymorphism MICA5.1 identifies extreme risk of progression to overt adrenal insufficiency among 21-hydroxylase antibody-positive patients with type 1 diabetes. J Clin Endocrinol Metab 94:4517–4523
14. O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266
15. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101
16. Raum D, Awdeh Z, Yunis EJ, Alper CA, Gabbay KH (1984) Extended major histocompatibility complex haplotypes in type I diabetes mellitus. J Clin Invest 74:449–454
17. Thomson G (1995) Mapping disease genes: family-based association studies. Am J Hum Genet 57:487–498
18. Baschal EE, Aly TA, Jasinski JM et al (2009) The frequent and conserved DR3-B8-A1 extended haplotype confers less diabetes risk than other DR3 haplotypes. Diabetes Obes Metab 11(Suppl 1):25–30
19. Rubinstein P, Walker M, Carpenter C et al (1981) Genetics of HLA–disease associations: the use of the haplotype relative risk

(HRR) and the "haplo-delta" (Dh) estimates in juvenile diabetes from three racial groups. Hum Immunol 3:384
20. Kaas Q, Ruiz M, Lefranc MP (2004) IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. Nucleic Acids Res 32: D208–D210
21. Baschal EE, Aly TA, Jasinski JM et al (2009) Defining multiple common "completely" conserved major histocompatibility complex SNP haplotypes. Clin Immunol 132:203–214
22. McKenzie LM, Pecon-Slattery J, Carrington M, O'Brien SJ (1999) Taxonomic hierarchy of HLA class I allele sequences. Genes Immun 1:120–129
23. Alper CA, Larsen CE, Dubey DP, Awdeh ZL, Fici DA, Yunis EJ (2006) The haplotype structure of the human major histocompatibility complex. Hum Immunol 67:73–84
24. Lipponen K, Gombos Z, Kiviniemi M et al (2010) Effect of HLA class I and class II alleles on progression from autoantibody positivity to overt type 1 diabetes in children with risk-associated class II genotypes. Diabetes 59:3253–3256
25. Noble JA, Valdes AM, Varney MD et al (2010) HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. Diabetes 59:2972–2979
26. Parham P, Lawlor DA, Lomen CE, Ennis PD (1989) Diversity and diversification of HLA-A, B, C alleles. J Immunol 142:3937–3950
27. Vargas-Alarcon G, Gomez-Casado E, Martinez-Laso J et al (1997) Differences in intron 2 sequences between *B*39061* and *B*39062* in Amerindians: comparison with those of *B*3901*, *B*5101*, and *B*52012* alleles. Immunogenetics 45:436–439
28. Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP (2007) Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8:762–775
29. Nejentsev S, Reijonen H, Adojaan B et al (1997) The effect of HLA-B allele on the IDDM risk defined by DRB1*04 subtypes and DQB1*0302. Diabetes 46:1888–1892
30. Gombos Z, Wachowicz J, Veijola R et al (2006) Human leukocyte antigen non-class II determinants for type 1 diabetes in the Finnish population. Hum Immunol 67:714–721